Deep Reinforcement Learning for BBU Placement and Routing in C-RAN

Zhengguang Gao[1], Jiawei Zhang[1], Shuangyi Yan[2], Yuming Xiao[1], Dimitra Simeonidou[2], Yuefeng Ji[1]
1.State Key Lab of Information Photonics and Optical Communications, BUPT, Beijing, 100876, China
2. High Performance Networks Group, University of Bristol, UK
E-mail address: (gaozg, zjw, yumingxiao, jyf)@bupt.edu.cn; (shuangyi.yan, Dimitra.Simeonidou)@bristol.ac.uk

Abstract: The paper proposed a deep reinforcement learning (DRL) based policy for BBU placement and routing in C-RAN. The simulation results show DRL-based policy reaches the near-optimal performance with a significantly reduced computing time.
OCIS codes: (060.4256) Network optimization; (060.4251) Networks, assignment and routing algorithms.

1. Introduction

Cloud-radio access networks (C-RAN) as a promising 5G mobile network architecture [1], can significantly decrease CapEx/OpEx for the high consolidation of base-band unit (BBU) and allow deployment of advanced technologies such as coordinated multipoint (CoMP) transmission/reception. To meet the severe requirements of mobile fronthaul (MFH) bandwidth in C-RAN, optical networks have been introduced to support C-RAN architecture. An appropriate strategy for BBU allocation and routing is required to reduce the required MFH bandwidth and satisfy strict latency requirement [2]. The strategy needs to handle the following challenges: 1) the place where BBU is deployed under a given task request in order to reduce mobile fronthaul bandwidth while maintaining the high consolidation of BBU; 2) the algorithm to find the shortest path to reduce transport delay. Generally, the strategy can be formulated as a network planning problem which can be solved by Optimal Method, or heuristic algorithm extended from Optimal Method [3]. The optimal method such as integer linear programming, can provide the best strategy with considerable number of iterations. To achieve the reasonable performance, the complexity and the required computing resource of the optimal method prohibit the real-time deployment. On the other hand, heuristic algorithms may satisfy the speed for real-time processing, but it can't ensure that a satisfactory strategy can be obtained.

Recently, deep reinforcement learning (DRL) has attracted much attention as an effective method to handle complicated problems such as playing computer games and autopilot. In particular, it can achieve the optimal mapping from state space into action space by deep neural network (DNN). And reinforcement learning has been used in resource allocation and slice Admission for 5G optical networks. In [4], a deep reinforcement learning based routing, modulation and spectrum assignment (RMSA) mechanism is proposed in elastic optical networking. In [5], a slice admission policy based on deep Q-network is studied to maximize the infrastructure provider's profit. Deep learning can perform end-to-end training and abstract a complex multi-layered model with strong expressive power, it is believed that reinforcement learning algorithms based DNN can solve very complex decision problems.

Inspired by this, we proposed a BBU placement and routing policy based on deep reinforcement learning in C-RAN to improve network resource utilization and to reduce network latency for 5G applications. The simulation results show that the performance of the proposed DRL-based algorithm far exceeds first-fit algorithm, and it can achieve about 98% performance of optimal benchmark by integer linear programming (ILP) while significantly decreasing the complexity for the computation. For example, the computation time is dropped from 28 seconds by ILP to less than 0.06 second by DRL-based algorithm in a 96-task sequence.

2. RAN architecture and DRL-based policy

2.1. RAN architecture

Fig. 1. (a) demonstrates the typical architecture of C-RAN proposed in this paper. Each central office (CO) aggregated from multiple active antenna units (AAUs) is connected by optical networks, which constitutes a converged wireless and optical network. Each CO in optical network can hold BBU to do the baseband processing, and the processed data is transported to date center (DC) by optical networks. Therefore, a completed request experiences two phases: mobile fronthaul from RRH to BBU and mobile backhaul from BBU to DC.

Based on the fact that the fronthaul will require much more bandwidth compared to backhaul for it transmits the raw data derived from multiple AAUs, we propose the policy to meet the huge bandwidth requirement and ultra-low latency for C-RAN architecture from the following three concepts: 1) we choose the active CO which has held the BBU to ensure the high consolidation of BBU if the computing resources are sufficient for a given request. 2) we

choose the proximal CO to the AAU with the request to avoid the fronthaul. 3) we choose the shortest path to decrease the latency. These concepts are interdependent and mutually restrictive. For example, the proximal CO to the AAU is selected for request 1 in Fig.1. (a), and this active CO is selected for request 2, but it can't ensure the shortest path. Therefore, we devise an objective function $f(N,B,T)=w\times N+f\times B+h\times T$ to balance these three points. The parameter N is the required number of COs which has held BBU, (B,T) are bandwidth and latency respectively, and (w, f, h) are the weights representing the priority for its corresponding factors. Therefore we can formulate these as an optimal problem and minimize the objective function $f(N,B,T)$ with some constraint considering the restrictions for computing resource, bandwidth and latency. However, the optimal method spends much time to search the satisfied result when the request sequence is large. So here we propose DRL-based algorithm for BBU placement and path selection. Integer linear programming and first-fit algorithm are also deployed as baseline algorithms to compare the effect of proposed DRL-based policy.

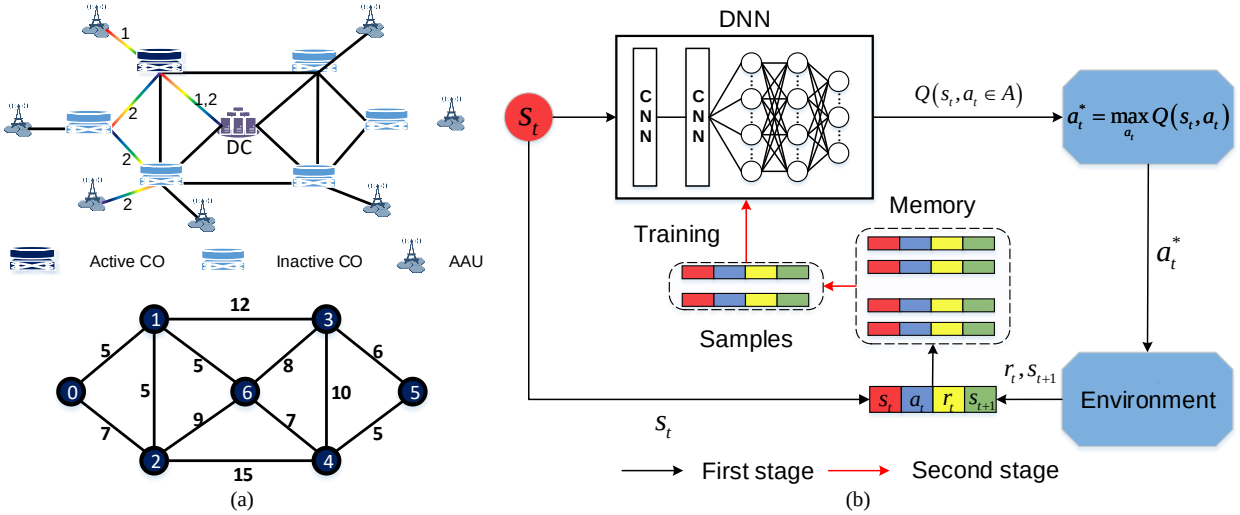


Fig. 1. (a) The typical architecture of C-RAN; (b) The proposed Schematic of DRL-based policy.

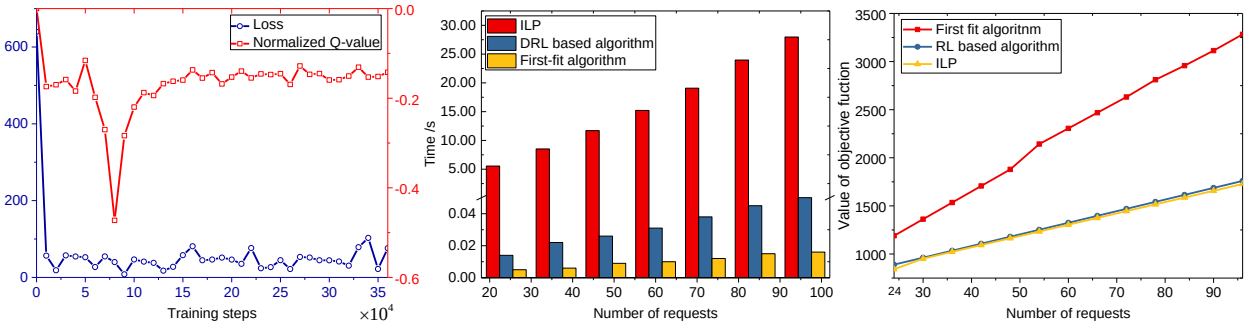
2.2. DRL-based algorithm

The structure of DRL-based algorithm is presented in Fig. 1. (b). DNN is built to extract complex mappings from the network states s_t to the output of corresponding action value $Q^{\pi_t}(s_t,a_t)$. And $Q^{\pi_t}(s_t,a_t)$ denotes the suitability of BBU allocation and path selection for the given optical network status s_t. In order to extract the state of optical network better, two convolutional neural networks with 5 convolution kernels are built at first, then two fully connected layers with tanh activation function are followed, where 245 and 100 hidden neurons are deployed. The output of DNN are 51 neurons with no activation function, which estimates the value of actions that includes all possible combination of BBU placement and path selection. The main concept of proposed DRL based algorithm includes the following two stages.

The first stage is to generate the training data. During the *t*-th time frame in fig.1. (b), current network state s_t comprised of the network's bandwidth, link's latency and node's computing resource is taken into DNN, then DNN outputs the value of all the possible actions $Q^{\pi_t}(s_t, a_t \in A)$ from the action space A under the current policy π_t, which is parameterized by DNN's parameters θ_t. Subsequently we select the action a_t^* with the largest value to interact with the environment and observe the next state s_{t+1} as well as the reward r_t which defined as $r(s_t,a_t)=-(w\times x_t+f\times b_t+h\times l_t)$, where $x_t\in\{0,1\}$ is an indicator variable. $x_t=1$ denotes one of inactive CO is selected to hold BBU while $x_t=0$ denotes the active CO is selected. b_t and l_t denote the bandwidth and latency for the current request. After this interaction, the data pair (s_t, r_t, a_t, s_{t+1}) is saved in the memory.

The second stage is to update the parameters of DNN. The core theory for it is Bellman equation of optimal Q-function as $Q^{\pi_t}(s_t,a_t)=r(s_t,a_t)+\gamma\max_{a_{t+1}}Q^{\pi_t}(s_{t+1},a_{t+1})$ where γ is a discount factor. Here we use DNN with parameters θ_t to fit the Q-function as $Q^{\pi_t}(s_t,a_t)\approx Q^{\pi_t}(s_t,a_t,\theta_t)$, $Q^{\pi_t}(s_{t+1},a_{t+1})\approx Q^{\pi_t}(s_{t+1},a_{t+1},\theta_t)$. So the optimal fit should be consistent with the Bellman equation as $Q^{\pi_t}(s_t,a_t,\theta_t)=r(s_t,a_t)+\gamma\max_{a_{t+1}}Q^{\pi_t}(s_{t+1},a_{t+1},\theta_t)$. Based on this, we

can minimize Bellman error $\xi = 0.5 * \mathrm{E}_{(s_t, a_t)}[Q^{\pi_t}(s_t, a_t, \theta_t) - r(s_t, a_t) - \gamma \max_{a_{t+1}} Q^{\pi_t}(s_{t+1}, a_{t+1}, \theta_t)]$ to update DNN parameters by gradient descent method. If Bellman error converges after updating the DNN parameters, the value of actions is fit well by DNN. So we can conclude DRL-based algorithm as: 1) we collect training data pairs (s_t, r_t, a_t, s_{t+1}) by interacting with the environment under the current policy π_t; 2) a random mini-batch of data (s_i, r_i, a_i, s_{i+1}) is sampled to update DNN parameters by $\theta_{t+1} \leftarrow \theta_t - \alpha \sum_i d\xi_i / d\theta_t$; 3) we use the new policy π_{t+1} to interact with the environment to collect new data to update the memory and repeat the first two steps until the algorithm converges.



(a) (b) (c)

Fig. 2. (a) The convergence property of proposed algorithm; (b) The computation time of three algorithms; (c) The algorithm performances.

3. Simulation setup and discussion

The performance of DRL-based algorithm for BBU placement and routing in C-RAN is presented in Fig. 2. The length of optical links is shown in network topology of Fig.1. (a). The bandwidth of each link is 25 Gbps, the total computing resource for each CO is 20000 GOPS. For a given request, the bandwidth requirement for each link in mobile fronthaul is 2.2948 Gbps while 0.2432 Gbps is needed for mobile backhaul [6]; The computing resource for CO to hold BBU is 1200 GOPS each time, and the maximum allowed fronthaul latency is 10 km. In the simulation, DRL-based policy is achieved after 20000 episodes. The max number of requests allowed for each episode depends on the bandwidth and computing resource for the links and COs in optical network. In this paper, this number of request allowed is 96. The weights for objective function and reward are set as (w, f, h)=(100,10,1).

Fig. 2. (a) presents the trend of fitting loss of DNN and normalized Q-value against the training steps. The loss first drops rapidly as the training begins, then gradually converges a small value. And the normalized Q-value first decreases rapidly for random exploration strategy decided by initial parameters of DNN, then increases quickly for continuous improving strategy with the training of DNN, finally approach to 0, the theoretical maximum value. We see that the strategy characterized by the output of DNN finally converges from the trend of loss and normalized Q-value. Fig. 2. (b) presents the computing time to generate strategy for the request sequence by ILP, first fit and DRL-based algorithms. The simulation results show that the proposed algorithm decreases more than 2 orders of magnitude compared with ILP. For example, the computing time decreases from 28 seconds by ILP to less than 0.06 second by DRL-based algorithm in a 96-task sequence. Fig. 2. (c) shows the value of objective function representing the resource consumed for these policies. We see that DRL-based algorithm significantly outperforms first-fit algorithm. Furthermore, the proposed algorithm has reached the near-optimal performance by ILP.

4. Conclusion

This work presents a DRL-based policy for BBU placement and routing to reduce the cost of required resource for the given requests. The results prove that the proposed algorithm outputs first-fit algorithm and reaches near-optimal performance achieved by ILP while reducing the computing time significantly.

Acknowledgements: This work was supported by the National Science and Technology Major Project (No. 2017ZX03001016), National Nature Science Foundation of China Projects (No. 61771073), the Fund of State Key Laboratory of Information Photonics and Optical Communications (Beijing University of Posts and Telecommunications), P. R. China. No. IPOC2017ZT09.

References

[1] A. Checko et al., "Cloud RAN for Mobile Networks – A Technology Overview," Commun. Surveys Tuts., 17, 405-426, 2015.
[2] F. Musumeci, et al., "Optimal BBU placement for 5G C-RAN deployment over WDM aggregation networks," JLT 34.8 (2016): 1963-1970.
[3] Yao Li, et al., "Joint Optimization of BBU Pool Allocation and Selection for C-RAN Networks." OFC, 2018.
[4] X. Chen, et al., "Deep-RMSA. " OFC, 2018.
[5] M. R. Raza, et al., "A Slice Admission Policy Based on Reinforcement Learning fora 5G Flexible RAN," ECOC, 2018.
[6] Desset, et al. "Flexible power modeling of LTE base stations." (WCNC). 2012.